No Silver Bullet

Essence and Accidents of Software Engineering

Frederick P. Brooks

1986

Abstract

All software construction involves *essential* tasks, the fashioning of the complex conceptual structures that compose the abstract software entity, and *accidental* tasks, the representation of these abstract entities in programming languages and the mapping of these onto machine languages within space and speed constraints. Most of the big past gains in software productivity have come from removing artificial barriers that have made the accidental tasks inordinately hard, such as severe hardware constraints, awkward programming languages, lack of machine time. How much of what software engineers now do is still devoted to the accidental, as opposed to the essential? Unless it is more than 9/10 of all effort, shrinking all the accidental activities to zero time will not give an order of magnitude improvement.

Therefore it appears that the time has come to address the essential parts of the software task, those concerned with fashioning abstract conceptual structures of great complexity. I suggest:

- exploiting the mass market to avoid constructing what can be bought.
- using rapid prototyping as part of a planned iteration in establishing software requirements.
- growing software organically, adding more and more function to systems as they are run, used, and tested.
- identifying and developing the great conceptual designers of the rising generation.

Introduction

Of all the monsters who fill the nightmares of our folklore, none terrify more than werewolves, because they transform unexpectedly from the familiar into horrors. For these, one seeks bullets of silver that can magically lay them to rest.

The familiar software project has something of this character (at least as seen by the non-technical manager), usually innocent and straightforward, but capable of becoming a monster of missed schedules, blown budgets, and flawed products. So we hear desperate cries for a silver bullet, something to make software costs drop as rapidly as computer hardware costs do.

But, as we look to the horizon of a decade hence, we see no silver bullet. There is no single development, in either technology or management technique, which by itself promises even one order of magnitude improvement in productivity, in reliability, in simplicity. In this paper we shall try to see why, by examining both the nature of the software problem and the properties of the bullets proposed.

Skepticism is not pessimism, however. Although we see no startling breakthroughs, and indeed, believe such to be inconsistent with the nature of software, many encouraging innovations are under way. A disciplined, consistent effort to develop, propagate, and exploit them should indeed yield an order-of-magnitude improvement. There is no royal road, but there is a road.

The first step toward the management of disease was replacement of demon theories and humours theories by the germ theory. That very step, the beginning of hope, in itself dashed all hopes of magical solutions. It told workers that progress would be made stepwise, at great effort, and that a persistent, unremitting care would have to be paid to a discipline of cleanliness. So it is with software engineering today.

Does it have to be Hard? – Essential Difficulties

Not only are there no silver bullets now in view, the very nature of software makes it unlikely that there will be any — no inventions that will do for software productivity, reliability, and simplicity what electronics, transistors, large-scale integration did for computer hardware. We cannot expect ever to see two-fold gains every two years.

First, one must observe that the anomaly is not that software progress is so slow, but that computer hardware progress is so fast. No other technology since civilization began has seen six orders of magnitude price-performance gain in 30 years. In no other technology can one choose to take the gain in either improved performance or in reduced costs. These gains flow from the transformation of computer manufacture from an assembly industry into a process industry.

Second, to see what rate of progress one can expect in software technology, let us examine its difficulties. Following Aristotle, I divide them into *essence*, the difficulties inherent in the nature of the software, and *accidents*, those difficulties which today attend its production but which are not inherent.

The accidents I discuss in the next section. First let us consider the essence.

The essence of a software entity is a construct of interlocking concepts: data sets, relationships among data items, algorithms, and invocations of functions. This essence is abstract, in that the conceptual construct is the same under many different representations. It is nonetheless highly precise and richly detailed.

I believe the hard part of building software to be the specification, design, and testing of this conceptual construct, not the labor of representing it and testing the fidelity of the representation. We still make syntax errors, to be sure; but they are fuzz compared to the conceptual errors in most systems.

If this is true, building software will always be hard. There is inherently no silver bullet.

Let us consider the inherent properties of this irreducible essence of modern software systems: complexity, conformity, changeability, and invisibility.

Complexity

Software entities are more complex for their size than perhaps any other human construct, because no two parts are alike (at least above the statement level). If they are, we make the two similar parts into one, a subroutine, open or closed. In this respect software systems differ profoundly from computers, buildings, or automobiles, where repeated elements abound.

Digital computers are themselves more complex than most things people build; they have very large numbers of states. This makes conceiving, describing, and testing them hard. Software systems have orders of magnitude more states than computers do.

Likewise, a scaling-up of a software entity is not merely a repetition of the same elements in larger size, it is necessarily an increase in the number of different elements. In most cases, the elements interact with each other in some non-linear fashion, and the complexity of the whole increases much more than linearly.

The complexity of software is an essential property, not an accidental one. Hence descriptions of a software entity that abstract away its complexity often abstract away its essence. Mathematics and the physical sciences for three centuries made great strides by constructing simplified models of complex phenomena, deriving properties from the models, and verifying those properties experimentally. This worked because the complexities ignored in the models were not the essential properties of the phenomena. It does not work when the complexities are the essence.

Many of the classical problems of developing software products derive from this essential complexity and its non-linear increases with size. From the complexity comes the difficulty of communication among team members, which leads to product flaws, cost overruns, schedule delays. From the complexity comes the difficulty of enumerating, much less understanding, all the possible states of the program, and from that comes the unreliability. From the complexity of the functions comes the difficulty of invoking those functions, which makes programs hard to use. From complexity of structure comes the difficulty of extending programs to new functions without creating side effects. From complexity of structure come the unvisualized states that constitute security trapdoors.

Not only technical problems, but management problems as well come from the complexity. It makes overview hard, thus impeding conceptual integrity. It makes it hard to find and control all the loose ends. It creates the tremendous learning and understanding burden that makes personnel turnover a disaster.

2.2 Conformity

Software people are not alone in facing complexity. Physics deals with terribly complex objects even at the “fundamental” particle level. The physicist labors on, however, in a firm faith that there are unifying principles to be found, whether in quarks or in unified field theories. Einstein repeatedly argued that there must be simplified explanations of nature, because God is not capricious or arbitrary.

No such faith comforts the software engineer. Much of the complexity he must master is arbitrary complexity, forced without rhyme or reason by the many human institutions and systems to which his interfaces must conform. These differ from interface to interface, and from time to time, not because of necessity but only because they were designed by different people, rather than by God.

In many cases the software must conform because it has most recently come to the scene. In others it must conform because it is perceived as the most conformable. But in all cases, much complexity comes from conformation to other interfaces; this cannot be simplified out by any redesign of the software alone.

2.3 Changeability

The software entity is constantly subject to pressures for change. Of course, so are buildings, cars, computers. But manufactured things are infrequently changed after manufacture; they are superseded by later models, or essential changes are incorporated in later serial-number copies of the same basic design. Call-backs of automobiles are really quite infrequent; field changes of computers somewhat less so. Both are much less frequent than modifications to fielded software.

Partly this is because the software in a system embodies its function, and the function is the part which most feels the pressures of change. Partly it is because software can be changed more easily — it is pure thought-stuff,

infinitely malleable. Buildings do in fact get changed, but the high costs of change, understood by all, serve to dampen the whims of the changers.

All successful software gets changed. Two processes are at work. As a software product is found to be useful, people try it in new cases at the edge of, or beyond, the original domain. The pressures for extended function come chiefly from users who like the basic function and invent new uses for it.

Second, successful software also survives beyond the normal life of the machine vehicle for which it is first written. If not new computers, then at least new disks, new displays, new printers come along; and the software must be conformed to its new vehicles of opportunity.

In short, the software product is embedded in a cultural matrix of applications, users, laws, and machine vehicles. These all change continually, and their changes inexorably force change upon the software product.

2.4 Invisibility

Software is invisible and unvisualizable. Geometric abstractions are powerful tools. The floor plan of a building helps both architect and client evaluate spaces, traffic flows, views. Contradictions become obvious, omissions can be caught. Scale drawings of mechanical parts and stick-figure models of molecules, although abstractions, serve the same purpose. A geometric reality is captured in a geometric abstraction.

The reality of software is not inherently embedded in space. Hence it has no ready geometric representation in the way that land has maps, silicon chips have diagrams, computers have connectivity schematics. As soon as we attempt to diagram software structure, we find it to constitute not one, but several, general directed graphs, superimposed one upon another. The several graphs may represent the flow of control, the flow of data, patterns of dependency, time sequence, name-space relationships. These are usually not even planar, much less hierarchical. Indeed, one of the ways of establishing conceptual control over such structure is to enforce link cutting until one or more of the graphs becomes hierarchical[1].

In spite of progress in restricting and simplifying the structures of software, they remain inherently unvisualizable, thus depriving the mind of some of its most powerful conceptual tools. This lack not only impedes the process of design within one mind, it severely hinders communication among minds.

1. Parnas, D.L., Designing Software for Ease of Extension and Contraction, *IEEE Trans. on SE*, vol. 5, no. 2, March, 1979, 128-138.

Past breakthroughs solved accidental difficulties

If we examine the three steps in software technology that have been most fruitful in the past, we discover that each attacked a different major difficulty in building software, but they have been the accidental, not the essential, difficulties. We can also see the natural limits to the extrapolation of each such attack.

High-Level Languages

Surely the most powerful stroke for software productivity, reliability, and simplicity has been the progressive use of high-level languages for programming. Most observers credit that development with at least a factor of five in productivity, and with concomitant gains in reliability, simplicity, and comprehensibility.

What does a high-level language accomplish? It frees a program from much of its accidental complexity. An abstract program consists of conceptual constructs: operations, data-types, sequences, and communication. The concrete machine program is concerned with bits, registers, conditions, branches, channels, disks, and such. To the extent that the high-level language embodies the constructs one wants in the abstract program and avoids all lower ones, it eliminates a whole level of complexity that was never inherent in the program at all.

The most a high-level language can do is to furnish all the constructs the programmer imagines in the abstract program. To be sure, the level of our sophistication in thinking about data structures, data types, and operations is steadily rising, but at an ever-decreasing rate. And language development approaches closer and closer to the sophistication of users.

Moreover, at some point the elaboration of a high-level language becomes a burden that increases, not reduces, the intellectual task of the user who rarely uses the esoteric constructs.

Time-Sharing

Most observers credit time-sharing with a major improvement in the productivity of programmers and in the quality of their product, although not so large as that brought by high-level languages.

Time-sharing attacks a quite different difficulty. Time-sharing preserves immediacy, and hence enables one to maintain an overview of complexity. The slow turnaround of batch programming means that one inevitably forgets the minutae, if not the very thrust, of what he was thinking when he stopped programming and called for compilation and execution. This interruption of consciousness is costly in time, for one must refresh. The most serious effect may well be the decay of grasp of all that is going on in a complex system.

Slow turn-around, like machine-language complexities, is an accidental rather than an essential difficulty of the software process. The limits of the contribution of time-sharing derive directly. The principal effect is to shorten system response time. As it goes to zero, at some point it passes the human threshold of noticeability, about 100 milliseconds. Beyond that no benefits are to be expected.

Unified Programming Environments

Unix and Interlisp, the first integrated programming environments to come into widespread use, are perceived to have improved productivity by integral factors. Why?

They attack the accidental difficulties of using programs *together*, by providing integrated libraries, unified file formats, and pipes and filters. As a result, conceptual structures that in principle could always call, feed, and use one another can indeed easily do so in practice.

This breakthrough in turn stimulated the development of whole tool benches, since each new tool could be applied to any programs using the standard formats.

Because of these successes, environments are the subject of much of today's software engineering research. We will look at their promise and limitations in the next section.

Hopes for the silver

Now let us consider the technical developments that are most often advanced as potential silver bullets. What problems do they address? Are they the problems of essence, or are they remainders of our accidental difficulties? Do they offer revolutionary advances, or incremental ones?

Ada and other High-Level Language Advances

One of the most touted recent developments is the programming language Ada, a general-purpose high-level language of the ’80’s. Ada indeed reflects not only evolutionary improvements in language concepts, but embodies features to encourage modern design and modularization concepts. Perhaps the Ada philosophy is more of an advance than the Ada language, for it is the philosophy of modularization, of abstract data types, of hierarchical structuring. Ada is perhaps over-rich, the natural product of the process by which requirements were laid on its design. That is not fatal, for subset working vocabularies can solve the learning problem, and hardware advances will give us the cheap MIPS to pay for the compiling costs. Advancing the structuring of software systems is indeed a very good use for the increased MIPS our dollars will buy. Operating systems, loudly decried in the ’60’s for their memory and cycle costs, have proved to be an excellent form in which to use some of the MIPS and cheap memory bytes of the past hardware surge.

Nevertheless, Ada will not prove to be the silver bullet that slays the software productivity monster. It is, after all, just another high-level language, and the biggest payoff from such languages came from the first transition, up from the accidental complexities of the machine into the more abstract statement of step-by-step solutions. Once those accidents have been removed, the remaining ones are smaller, and the payoff from their removal will surely be less.

I predict that a decade from now, when the effectiveness of Ada is assessed, it will be seen to have made a substantial difference, but not because of any particular language feature, nor indeed of all of them combined. Neither will the new Ada environments prove to be the cause of the improvements. Ada’s

greatest contribution will be that switching to it occasioned training programmers in modern software design techniques.

Object-Oriented Programming

Many students of the art hold out more hope for object-oriented programming than for any of the other technical fads of the day[1]. I am among them. Mark Sherman of Dartmouth notes that one must be careful to distinguish two separate ideas that go under that name: abstract data types and hierarchical types, also called classes. The concept of the abstract data type is that an object's type should be defined by a name, a set of proper values, and a set of proper operations, rather than its storage structure, which should be hidden. Examples are Ada packages (with private types) or Modula's modules.

Hierarchical types, such as Simula-67's classes, allow one to define general interfaces that can be further refined by providing subordinate types. The two concepts are orthogonal — one may have hierarchies without hiding and hiding without hierarchies. Both concepts represent real advances in the art of building software.

Each removes one more accidental difficulty from the process, allowing the designer to express the essence of his design without having to express large amounts of syntactic material that add no new information content. For both abstract types and hierarchical types, the result is to remove a higher-order sort of accidental difficulty and allow a higher-order expression of design.

Nevertheless, such advances can do no more than to remove all the accidental difficulties from the expression of the design. The complexity of the design itself is essential; and such attacks make no change whatever in that. An order-of-magnitude gain can be made by object-oriented programming only if the unnecessary underbrush of type specification remaining today in our programming language is itself responsible for nine-tenths of the work involved in designing a program product. I doubt it.

Artificial Intelligence

Many people expect advances in artificial intelligence to provide the revolutionary breakthrough that will give order-of-magnitude gains in software productivity and quality[2]. I do not. To see why, we must dissect what is meant by "artificial intelligence", and then see how it applies.

Parnas has clarified the terminological chaos[3]:

> Two quite different definitions of AI are in common use today. AI-1: The use of computers to solve problems that previously could only be solved by applying human intelligence. AI-2: The use of a specific set of programming techniques known as heuristic or rule-based programming. In this approach human experts are studied to determine what heuristics or rules of thumb they use in solving problems … The program is designed to solve a problem the way that humans seem to solve it.
>
> The first definition has a sliding meaning … Something can fit the definition of AI-1 today but, once we see how the program works and understand the problem, we will not think of it as AI any more … Unfortunately I cannot identify a body of technology that is unique to this field … Most of the work is problem-specific, and some abstraction or creativity is required to see how to transfer it.

I agree completely with this critique. The techniques used for speech recognition seem to have little in common with those used for image recognition, and both are different from those used in expert systems. I have a hard time seeing how image recognition, for example, will make any appreciable difference in programming practice. The same is true of speech recognition. The hard thing about building software is deciding what one wants to say, not saying it. No facilitation of expression can give more than marginal gains.

Expert systems technology, AI-2, deserves a section of its own.

Expert Systems

The most advanced part of the artificial intelligence art, and the most widely applied, is the technology for building expert systems. Many software scientists are hard at work applying this technology to the software-building environment [4] and [5]. What is the concept, and what are the prospects?

An expert system is a program containing a generalized inference engine and a rule base, designed to take input data and assumptions and explore the logical consequences through the inferences derivable from the rule base, yielding conclusions and advice, and offering to explain its results by retracing its reasoning for the user. The inference engines typically can deal with fuzzy or probabilistic data and rules in addition to purely deterministic logic.

Such systems offer some clear advantages over programmed algorithms for arriving at the same solutions to the same problems:

- Inference engine technology is developed in an application-independent way, and then applied to many uses. One can justify much more effort on the inference engines. Indeed, that technology is well advanced.
- The changeable parts of the application-peculiar materials are encoded in the rule base in a uniform fashion, and tools are provided for developing, changing, testing, and documenting the rule base. This regularizes much of the complexity of the application itself.

Edward Feigenbaum, says that the power of such systems does not come from ever-fancier inference mechanisms, but rather from ever-richer knowledge bases that reflect the real world more accurately. I believe the most important advance offered by the technology is the separation of the application complexity from the program itself.

How can this be applied to the software task? In many ways: suggesting interface rules, advising on testing strategies, remembering bug-type frequencies, offering optimization hints, etc.

Consider an imaginary testing advisor, for example. In its most rudimentary form, the diagnostic expert system is very like a pilot's checklist, fundamentally offering suggestions as to possible causes of difficulty. As the rule base is developed, the suggestions become more specific, taking more sophisticated account of the trouble symptoms reported. One can visualize a debugging assistant which offers very generalized suggestions at first, but as more and more system structure is embodied in the rule base, becomes more and more particular in the hypotheses it generates and the tests it recommends. Such an expert system may depart most radically from the conventional ones in that its rule base should probably be hierarchically modularized in the same way the corresponding software product is, so that as the product is modularly modified, the diagnostic rule base can be modularly modified as well.

The work required to generate the diagnostic rules is work that will have to be done anyway in generating the set of test cases for the modules and for the system. If it is done in a suitably general manner, with a uniform structure for rules and a good inference engine available, it may actually reduce the total labor of generating bring-up test cases, as well as helping in life-long maintenance and modification testing. In the same way, one can postulate other, probably many and probably simple, advisors for the other parts of the software construction task.

Many difficulties stand in the way of the early realization of useful expert advisors to the program developer. A crucial part of our imaginary scenario is the development of easy ways to get from program structure specification to the automatic or semiautomatic generation of diagnostic rules. Even more difficult and important is the two-fold task of knowledge-acquisition: finding articulate, self-analytical experts who know *why* they do things; and developing efficient techniques for extracting what they know and distilling it into rule bases. The essential prerequisite for building an expert system is to have an expert.

The most powerful contribution of expert systems will surely be to put at the service of the inexperienced programmer the experience and accumulated wisdom of the best programmers. This is no small contribution. The gap between the best software engineering practice and the average practice is very wide — perhaps wider than in any other engineering discipline. A tool that disseminates good practice would be important.

"Automatic" Programming

For almost 40 years, people have been anticipating and writing about "automatic programming", the generation of a program for solving a problem from a statement of the problem specifications. Some today write as if they expected this technology to provide the next breakthrough[6].

Parnas[7] implies that the term is used for glamor and not semantic content, asserting,

> In short, automatic programming always has been a euphemism for programming with a higher-level language than was presently available to the programmer.

He argues, in essence, that in most cases it is the solution method, not the problem, whose specification has to be given.

One can find exceptions. The technique of building generators is very powerful, and it is routinely used to good advantage in programs for sorting. Some systems for integrating differential equations have also permitted direct specification of the problem, and the system assessed the parameters, chose from a library of methods of solution, and generated the programs.

These applications have very favorable properties:

- The problems are readily characterized by relatively few parameters.

- There are many known methods of solution to provide a library of alternatives.
- Extensive analysis has led to explicit rules for selecting solution techniques, given problem parameters.

It is hard to see how such techniques generalize to the wider world of the ordinary software system, where cases with such neat properties are the exception. It is hard even to imagine how this breakthrough in generalization could conceivably occur.

Graphical Programming

A favorite subject for Ph.D. dissertations in software engineering is graphical, or visual, programming, the application of computer graphics to software design [8] and [9]. Sometimes the promise of such an approach is postulated from the analogy with VLSI chip design, where computer graphics plays so fruitful a role. Sometimes the approach is justified by considering flowcharts as the ideal program design medium, and providing powerful facilities for constructing them.

Nothing even convincing, much less exciting, has yet emerged from such efforts. I am persuaded that nothing will.

In the first place, as I have argued elsewhere[10], the flowchart is a very poor abstraction of software structure. Indeed, it is best viewed as Burks, von Neumann, and Goldstine's attempt to provide a desperately needed high-level control language for their proposed computer. In the pitiful, multi-page, connection-boxed form to which the flowchart has today been elaborated, it has proved to be essentially useless as a design tool — programmers draw flowcharts after, not before, writing the programs they describe.

Secondly, the screens of today are too small, in pixels, to show both the scope and the resolution of any serious detailed software diagram. The so-called "desktop metaphor" of today's workstation is instead an "airplane-seat" metaphor. Anyone who has shuffled a lap full of papers while seated in coach between two portly passengers will recognize the difference — one can see only a very few things at once. The true desktop provides overview of and random access to, a score of pages. Moreover, when fits of creativity run strong, more than one programmer or writer has been known to abandon the desktop for the more spacious floor. The hardware technology will have to advance quite substantially before the scope of our scopes is sufficient to the software design task.

More fundamentally, as I have argued above, software is very difficult to visualize. Whether one diagrams control flow, variable scope nesting, variable cross-references, data flow, hierarchical data structures, or whatever, one feels only one dimension of the intricately interlocked software elephant. If one superimposes all the diagrams generated by the many relevant views, it is difficult to extract any global overview. The VLSI analogy is fundamentally misleading — a chip design is a layered two-dimensional object whose geometry reflects its essence. A software system is not.

Program Verification

Much of the effort in modern programming goes into testing and the repair of bugs. Is there perhaps a silver bullet to be found by eliminating the errors at the source, in the system design phase? Can both productivity and product reliability be radically enhanced by following the profoundly different strategy of proving designs correct before the immense effort is poured into implementing and testing them?

I do not believe we will find the magic here. Program verification is a very powerful concept, and it will be very important for such things as secure operating system kernels. The technology does not promise, however, to save labor. Verifications are so much work that only a few substantial programs have ever been verified.

Program verification does not mean error-proof programs. There is no magic here, either. Mathematical proofs also can be faulty. So whereas verification might reduce the program-testing load, it cannot eliminate it.

More seriously, even perfect program verification can only establish that a program meets its specification. The hardest part of the software task is arriving at a complete and consistent specification, and much of the essence of building a program is in fact the debugging of the specification.

Environments and Tools

How much more gain can be expected from the exploding researches into better programming environments? One’s instinctive reaction is that the big-payoff problems were the first attacked, and have been solved: hierarchical file systems, uniform file formats so as to have uniform program interfaces, and generalized tools. Language-specific smart editors are developments not yet widely used in practice, but the most they promise is freedom from

syntactic errors and simple semantic errors.

Perhaps the biggest gain yet to be realized in the programming environment is the use of integrated database systems to keep track of the myriads of details that must be recalled accurately by the individual programmer and kept current in a group of collaborators on a single system.

Surely this work is worthwhile, and surely it will bear some fruit in both productivity and reliability. But by its very nature, the return from now on must be marginal.

Workstations

What gains are to be expected for the software art from the certain and rapid increase in the power and memory capacity of the individual workstation? Well, how many MIPS can one use fruitfully? The composition and editing of programs and documents is fully supported by today's speeds. Compiling could stand a boost, but a factor of 10 in machine speed would surely leave think-time the dominant activity in the programmer's day. Indeed, it appears to be so now.

More powerful workstations we surely welcome. Magical enhancements from them we cannot expect.

1. Booch, G., Object-Oriented Design, *Software Engineering with Ada*, Benjamin/Cummings, Menlo Park, CA, 1983.

2. Mostow, J., editor, Special Issue on Artificial Intelligence and Software Engineering, *IEEE Trans. on SE*, vol. 11, no. 11, November, 1985.

3. Parnas, D.L., Software Aspects of Strategic Defense Systems, *American Scientist*, November, 1985.

4. Balzer, R., A 15-Year Perspective on Automatic Programming, in Mostow, Special Issue on Artificial Intelligence and Software Engineering, *IEEE Trans. on SE*, vol. 11, no. 11, November, 1985, 1257-1267.

5. Mostow, J., editor, Special Issue on Artificial Intelligence and Software Engineering, *IEEE Trans. on SE*, vol. 11, no. 11, November 1985.

6. Balzer, R., A 15-Year Perspective on Automatic Programming, in Mostow, Special Issue on Artificial Intelligence and Software Engineering, *IEEE Trans.*

on SE, vol. 11, no. 11, November, 1985, 1257-1267.

7. Parnas, D.L., Software Aspects of Strategic Defense Systems, *American Scientist*, November, 1985.

8. Graphton, R.B. & T. Ichikawa, eds., Special Issue on Visual Programming, *Computer*, vol. 18, no. 8, August, 1985.

9. Raeder, G., A Survey of Current Graphical Programming Techniques, in Grafton & Ichikawa, eds., Special Issue on Visual Programming, *Computer*, vol. 18, no. 8, August, 1985, 11-25.

10. Brooks, F.P., *The Mythical Man-Month*, Addison Wesley Publishing Co., New York, 1975, Chapter 14.

Promising attacks on the Conceptual Essence

Even though no technological breakthrough promises to give the sort of magical results with which we are so familiar in the hardware area, there is both an abundance of good work going on now, and the promise of steady, if unspectacular progress.

All of the technological attacks on the accidents of the software process are fundamentally limited by the productivity equation:

$$\text{time of task} = \Sigma i (\text{frequency})i \times (\text{time})i$$

If, as I believe, the conceptual components of the task are now taking most of the time, then no amount of activity on the task components that are merely the expression of the concepts can give large productivity gains.

Hence we must consider those attacks that address the essence of the software problem, the formulation of these complex conceptual structures. Fortunately, some of these are very promising.

Buy versus Build

The most radical possible solution for constructing software is not to construct it at all.

Every day this becomes easier, as more and more vendors offer more and better software products for a dizzying variety of applications. While we software engineers have labored on production methodology, the personal computer revolution has created not one, but many, mass markets for software. Every newstand carries monthly magazines which, sorted by machine type, advertise and review dozens of products at prices from a few dollars to a few hundred dollars. More specialized sources offer very powerful products for the workstation and other Unix markets. Even software tools and environments can be bought off-the-shelf. I have elsewhere

proposed a marketplace for individual modules.

Any such product is cheaper to buy than to build afresh. Even at a cost of one hundred thousand dollars, a purchased piece of software is costing only about as much as one programmer-year. And delivery is immediate! Immediate at least for products that really exist, products whose developer can refer the prospect to a happy user. Moreover, such products tend to be much better documented and somewhat better maintained than homegrown software.

The development of the mass market is, I believe, the most profound long-run trend in software engineering. The cost of software has always been development cost, not replication cost. Sharing that cost among even a few users radically cuts the per-user cost. Another way of looking at it is that the use of *n* copies of a software system effectively multiplies the productivity of its developers by *n*. That is an enhancement of the productivity of the discipline and of the nation.

The key issue, of course, is applicability. Can I use an available off-the-shelf package to do my task? A surprising thing has happened here. During the ’50’s and ’60’s, study after study showed that users would not use off-the-shelf packages for payroll, inventory control, accounts receivable, etc. The requirements were too specialized, the case-to-case variation too high. During the ’80’s, we find such packages in high demand and widespread use. What has changed?

Not really the packages. They may be somewhat more generalized and somewhat more customizable than formerly, but not much. Not really the applications, either. If anything, the business and scientific needs of today are more diverse, more complicated than those of twenty years ago.

The big change has been in the hardware/software cost ratio. The buyer of a two-million dollar machine in 1960 felt that he could afford $250,000 more for a customized payroll program, one that slipped easily and non-disruptively into the computer-hostile social environment. The buyer of a $50,000 dollar office machine today cannot conceivably afford a customized payroll program; so he adapts his payroll procedure to the packages available. Computers are now so commonplace, if not yet so beloved, that the adaptations are accepted as a matter of course.

There are dramatic exceptions to my argument that the generalization of the software packages has changed little over the years: electronic spreadsheets and simple database systems. These powerful tools, so obvious in retrospect and yet so late appearing, lend themselves to myriads of uses, some quite

unorthodox. Articles and even books now abound on how to tackle unexpected tasks with the spreadsheet. Large numbers of applications that would formerly have been written as custom programs in Cobol or Report Program Generator are now routinely done with these tools.

Many users now operate their own computers day in and day out on varied applications without ever writing a program. Indeed, many of these users cannot write new programs for their machines, but they are nevertheless adept at solving new problems with them.

I believe the single most powerful software productivity strategy for many organizations today is to equip the computer-naive intellectual workers on the firing line with personal computers and good generalized writing, drawing, file, and spreadsheet programs, and turn them loose. The same strategy, with generalized mathematical and statistical packages and some simple programming capabilities, will also work for hundreds of laboratory scientists.

Requirements Refinement and Rapid Prototyping

The hardest single part of building a software system is deciding precisely what to build. No other part of the conceptual work is so difficult as establishing the detailed technical requirements, including all the interfaces to people, to machines, and to other software systems. No other part of the work so cripples the resulting system if done wrong. No other part is more difficult to rectify later.

Therefore the most important function that the software builder does for his client is the iterative extraction and refinement of the product requirements. For the truth is, the client does not know what he wants. He usually does not know what questions must be answered, and he almost never has thought of the problem in the detail that must be specified. Even the simple answer — “Make the new software system work like our old manual information-processing system” — is in fact too simple. One never wants exactly that. Complex software systems are, moreover, things that act, that move, that work. The dynamics of that action are hard to imagine. So in planning any software activity, it is necessary to allow for an extensive iteration between the client and the designer as part of the system definition.

I would go a step further and assert that it is really impossible for a client, even working with a software engineer, to specify completely, precisely, and

correctly the exact requirements of a modern software product before having built and tried some versions of the product he is specifying.

Therefore one of the most promising of the current technological efforts, and one which attacks the essence, not the accidents, of the software problem, is the development of approaches and tools for rapid prototyping of systems as part of the iterative specification of requirements.

A prototype software system is one which simulates the important interfaces and performs the main functions of the intended system, while not being necessarily bound by the same hardware speed, size, or cost constraints. Prototypes typically perform the mainline tasks of the application, but make no attempt to handle the exceptions, respond correctly to invalid inputs, abort cleanly, etc. The purpose of the prototype is to make real the conceptual structure specified, so that the client can test it for consistency and usability.

Much of present-day software acquisition procedures rests upon the assumption that one can specify a satisfactory system in advance, get bids for its construction, have it built, and install it. I think this assumption is fundamentally wrong, and that many software acquisition problems spring from that fallacy. Hence they cannot be fixed without fundamental revision, one which provides for iterative development and specification of prototypes and products.

Incremental Development - Grow, not Build, Software

I still remember the jolt I felt in 1958 when I first heard a friend talk about *building* a program, as opposed to *writing* one. In a flash he broadened my whole view of the software process. The metaphor shift was powerful, and accurate. Today we understand how like other building processes the construction of software is, and we freely use other elements of the metaphor, such *as specifications, assembly of components, and scaffolding*.

The building metaphor has outlived its usefulness. It is time to change again. If, as I believe, the conceptual structures we construct today are too complicated to be accurately specified in advance, and too complex to be built faultlessly, then we must take a radically different approach.

Let us turn to nature and study complexity in living things, instead of just the dead works of man. Here we find constructs whose complexities thrill us with

awe. The brain alone is intricate beyond mapping, powerful beyond imitation, rich in diversity, self-protecting, and self-renewing. The secret is that it is grown, not built.

So it must be with our software systems. Some years ago Harlan Mills proposed that any software system should be grown by incremental development[1]. That is, the system should first be made to run, even though it does nothing useful except call the proper set of dummy subprograms. Then, bit by bit it is fleshed out, with the subprograms in turn being developed into actions or calls to empty stubs in the level below.

I have seen most dramatic results since I began urging this technique on the project builders in my Software Engineering Laboratory class. Nothing in the past decade has so radically changed my own practice, or its effectiveness. The approach necessitates top-down design, for it is a top-down growing of the software. It allows easy backtracking. It lends itself to early prototypes. Each added function and new provision for more complex data or circumstances grows organically out of what is already there.

The morale effects are startling. Enthusiasm jumps when there is a running system, even a simple one. Efforts redouble when the first picture from a new graphics software system appears on the screen, even if it is only a rectangle. One always has, at every stage in the process, a working system. I find that teams can *grow* much more complex entities in four months than they can *build*.

The same benefits can be realized on large projects as on my small ones[2].

Great Designers

The central question in how to improve the software art centers, as it always has, on people.

We can get good designs by following good practices instead of poor ones. Good design practices can be taught. Programmers are among the most intelligent part of the population, so they can learn good practice. Thus a major thrust in the United States is to promulgate good modern practice. New curricula, new literature, new organizations such as the Software Engineering Institute, all have come into being in order to raise the level of our practice from poor to good. This is entirely proper.

Nevertheless, I do not believe we can make the next step upward in the same way. Whereas the difference between poor conceptual designs and good ones

may lie in the soundness of design method, the difference between good designs and great ones surely does not. Great designs come from great designers. Software construction is a *creative* process. Sound methodology can empower and liberate the creative mind; it cannot enflame or inspire the drudge.

The differences are not minor — it is rather like Salieri and Mozart. Study after study show that the very best designers produce structures that are faster, smaller, simpler, cleaner, and produced with less effort. The differences between the great and the average approach an order of magnitude.

A little retrospection shows that although many fine, useful software systems have been designed by committees and built by multipart projects, those software systems that have excited passionate fans are those which are the products of one or a few designing minds, great designers. Consider Unix, APL, Pascal, Modula, the Smalltalk interface, even Fortran; and contrast with Cobol, PL/I, Algol, MVS/370, and MS-DOS.

Hence, although I strongly support the technology transfer and curriculum development efforts now underway, I think the most important single effort we can mount is to develop ways to grow great designers.

No software organization can ignore this challenge. Good managers, scarce though they be, are no scarcer than good designers. Great designers and great managers are both very rare. Most organizations spend considerable effort in finding and cultivating the management prospects; I know of none that spends equal effort in finding and developing the great designers upon whom the technical excellence of the products will ultimately depend.

My first proposal is that each software organization must determine and proclaim that great designers are as important to its success as great managers are, and that they can be expected to be similarly nurtured and rewarded. Not only salary, but the perquisites of recognition — office size, furnishings, personal technical equipment, travel funds, staff support — must be fully equivalent.

How to grow great designers? Space does not permit a lengthy discussion, but some steps are obvious:

- Systematically identify top designers as early as possible. The best are often not the most experienced.
- Assign a career mentor to be responsible for the development of the prospect, and keep a careful career file.
- Devise and maintain a career development plan for each prospect,

including carefully selected apprenticeships with top designers, episodes of advanced formal education, and short courses, all interspersed with solo design and technical leadership assignments.
- Provide opportunities for growing designers to interact with and stimulate each other.

1. Mills, H.D., Top-Down Programming in Large Systems, in *Debugging Techniques in Large Systems*, R. Ruskin ed., Prentice-Hall, 1971.

2. Boehm, B.W., A Spiral Model of Software Development and Enhancement, TRW Technical Report, 1985.

Acknowledgements

I thank Gordon Bell, Bruce Buchanan, Rick Hayes-Roth, Robert Patrick, and, most especially, David Parnas for their insights and stimulating ideas, and Rebekah Bierly for technical production.